

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2003 年 12 月 31 日 (31.12.2003)

PCT

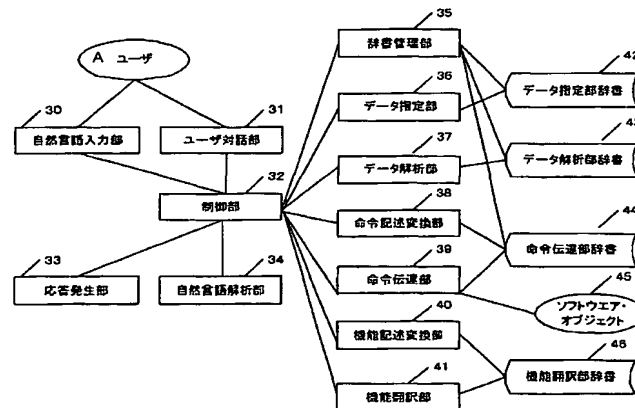
(10) 国際公開番号
WO 2004/001570 A1

- (51) 国際特許分類⁷: G06F 3/00, 17/27, G10L 13/08 (74) 代理人: 小林 良平 (KOBAYASI, Ryohei); 〒600-8091 京都府 京都市下京区 東洞院通四條下ル元恵王子町 3 7 豊元四條烏丸ビル 7 階 小林特許商標事務所 Kyoto (JP).
- (21) 国際出願番号: PCT/JP2003/002278
- (22) 国際出願日: 2003 年 2 月 27 日 (27.02.2003)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願2002-178303 2002 年 6 月 19 日 (19.06.2002) JP
- (71) 出願人 および
(72) 発明者: 荒木 修一 (ARAKI, Shuichi) [JP/JP]; 〒520-0822 滋賀県 大津市 秋葉台 3-8 Shiga (JP). 小倉 道雄 (OGURA, Michio) [JP/JP]; 〒603-8355 京都府 京都市北区 平野上柳町 1 1-7 2 Kyoto (JP).
- (81) 指定国 (国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) 指定国 (広域): ARIPO 特許 (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI

[続葉有]

(54) Title: METHOD FOR DESCRIBING EXISTING DATA BY A NATURAL LANGUAGE AND PROGRAM FOR THAT

(54) 発明の名称: 自然言語による既存データの記述方法及びそのためのプログラム



A...USER
 30...NATURAL LANGUAGE INPUT UNIT
 31...USER INTERACTION UNIT
 32...CONTROL UNIT
 33...RESPONSE GENERATION UNIT
 34...NATURAL LANGUAGE ANALYSIS UNIT
 35...DICTIONARY MANAGEMENT UNIT
 36...DATA SPECIFICATION UNIT
 37...DATA ANALYSIS UNIT
 38...INSTRUCTION DESCRIPTION CONVERSION UNIT
 39...INSTRUCTION TRANSMISSION UNIT
 40...FUNCTION DESCRIPTION CONVERSION UNIT
 41...FUNCTION TRANSLATION UNIT
 42...DATA SPECIFICATION UNIT DICTIONARY
 43...DATA ANALYSIS UNIT DICTIONARY
 44...INSTRUCTION TRANSMISSION UNIT DICTIONARY
 45...SOFTWARE OBJECT
 46...FUNCTION TRANSLATION UNIT DICTIONARY

(57) Abstract: Existing data created by certain software can easily be used by another software or for another object. This object can be achieved as follows. Existing data created by a specific software object is analyzed and converted into an instruction description string executed by the data by the software object. This instruction description string is converted into a function description expression using a

[続葉有]

BEST AVAILABLE COPY



特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

添付公開書類:

— 国際調査報告書

natural language which can be understood by a user. From this function description expression, meaningful information is acquired and added to the function description expression, thereby creating a request description expression. This request description expression is described by a natural language and includes all the information required for reproducing the original data.

(57) 要約: 本発明は、あるソフトウェアで作成した既存のデータを他のソフトウェアで利用したり、他の目的で利用したりすることを容易化することを目的として成されたものである。このような目的は、特定のソフトウェア・オブジェクトで作成された既存のデータを解析し、前記ソフトウェア・オブジェクトがこのデータを作成するために実行した命令記述列に変換し、この命令記述列をユーザに理解可能な自然言語による機能記述表現に変換し、この機能記述表現から意味的な情報を獲得し、前記機能記述表現に意味を付加した要求記述表現を生成することにより達成される。この要求記述表現は自然言語で記述されており、元のデータを再現するために必要な情報が全て含まれている。

明 細 書

自然言語による既存データの記述方法及びそのためのプログラム

技術分野

本発明は、コンピュータ上で動作するソフトウェア・オブジェクトにより作成された既存データを自然言語で記述する技術に関し、特に、自然言語で操作可能なソフトウェア・オブジェクトに入力すべき自然言語文字列を既存データから生成する技術に関する。本明細書においてソフトウェア・オブジェクトとは、パーソナル・コンピュータやマイコン制御機器等の電子機器を制御するためのオペレーティング・システム（OS）や、OS上で動作するアプリケーション・プログラムのことをいう。また、入力機器（キーボード、マイク、手書きタブレット等）から受け取った信号を処理して自然言語の文字列を生成し、その文字列を解析し、解析結果に基づいてソフトウェア・オブジェクトの操作命令を生成するシステムのことを、本明細書では自然言語インターフェースと呼ぶ。

背景技術

ソフトウェア・オブジェクトを自然言語で操作するための自然言語インターフェースに関する研究は従来より盛んに行われている。例として、特開平8-147096号公報に記載の手書き入力方法及び装置、特開平6-75692号公報に記載の情報処理装置、特開平6-131108号公報に記載の情報入力装置、特開平6-282566号公報に記載の情報入力装置等が挙げられる。これら従来の自然言語インターフェースは、ソフトウェア・オブジェクトに用意された内部機能を自然言語で呼び出すために用いられる。例えば、特開平6-75692には、ユーザが「大きくする」と手書き入力すると予め指定された文字列が倍角文字になるようなワードプロセッサが開示されている。また、特開平8-147096には、ユーザが「録画」と手書き入力すると録画処理が始まるような制御系を有するビデオデッキが開示されている。

従来の自然言語インターフェースは、例えばワードプロセッサ・プログラムやビデオデッキの制御プログラムといった個々のソフトウェア・オブジェクトに特

化したものばかりであって、あるソフトウェア・オブジェクトのために開発された自然言語インターフェースを他のソフトウェア・オブジェクトの操作に転用することは基本的に想定されていない。このため、ソフトウェア開発者は、あるソフトウェア・オブジェクト用の自然言語インターフェースを装備したい場合、専用の自然言語インターフェースを新たに開発するために相当な労力を費やさなければならない。

また、従来の自然言語インターフェースでは、ソフトウェア・オブジェクトに予め用意された内部機能と呼び出すための命令が入力されることが想定されている。従って、ユーザは、ソフトウェア・オブジェクトがどのような機能を備えており、それをどのような自然言語で呼び出せばよいかに関する情報（知識）を予め持ち合わせていなければならない。つまり、ユーザの要求に応える形でソフトウェア・オブジェクトが機能するというより、ソフトウェア・オブジェクトの機能に合わせてユーザが命令を出すのである。このような形態に留まる限り、自然言語によるソフトウェア・オブジェクトの操作は柔軟性を欠いたものとならざるを得ない。例えば、ユーザが「映画鑑賞会のお知らせを作成したい」と考え、その考えをそのまま言葉にして入力したとする。「映画鑑賞会のお知らせを作成したい」は、いわばユーザの要求、願望又は目的を表現したものであって、ソフトウェア・オブジェクトの何らかの機能を明示的に呼び出すための命令ではない。従来の自然言語インターフェースではこのような入力に対して適切な処理を実行することは不可能である。

上記課題に鑑み、本願出願人らは、複数の異なるソフトウェア・オブジェクトを統一的に操作できる汎用性と、ユーザの要求、願望、目的等の自然言語表現を入力として受け取ったときでも適切な処理ができる柔軟性を備えた自然言語インターフェースに関する発明を成した（特願2002-76319）。この発明では、入力された自然言語の文字列をユーザの要求の表現ととらえて文字列の解析を行い、その要求に応じた処理を行うのに最適なソフトウェア・オブジェクトを選択し、そのソフトウェア・オブジェクトに前記処理を行わせるための機能記述表現を中間的に生成し、その機能記述表現をOS又はプログラムにより実行可能な命令記述列に変換する。

一方、従来のソフトウェア・オブジェクトでは、ユーザの要求を実現するために生成するデータを、そのソフトウェア・オブジェクトだけが理解可能な専用の書式で保存していることが多い。ソフトウェア・オブジェクトが生成するデータには、文書や映像などのコンテンツ・データや、他のソフトウェア・オブジェクトやハードウェア機器などを制御するための制御シーケンス・データなどがある。しかし、その書式は特定のソフトウェア・オブジェクトだけが理解可能なため、データ自体の再利用性に乏しい。例えば、あるワードプロセッサで既に作成した文書データを、他のワードプロセッサで内容を編集して再利用したい場合がある。その場合、その文書データを特定のワードプロセッサが理解可能な専用の書式に変換するか、どのワードプロセッサでも理解可能な汎用の書式に変換することが必要になる。また、ワードプロセッサで作成した文書データを、他の種類のソフトウェア・オブジェクト、例えばメール・アプリケーションやプレゼンテーション・アプリケーションを使って他の用途に転用したい場合も、同様の書式の問題のため、煩雑なデータ変換作業が必要であったり、用途が限定されてしまうことが多い。

さらにまた、従来のソフトウェア・オブジェクトでは、ユーザの要求を実現するために生成するデータに、ユーザの要求を実現するための必要最低限の情報しか含まないことが一般的である。例えば、ワードプロセッサが生成する文書データには、文書のタイトルや作成日、作成者などの文字列が含まれているが、どの文字列がタイトルで、どの文字列が作成日なのかを特定する情報（いわゆる文字列の意味的情報）は一般に含まれていない。このため、ユーザが文書データの一部を修正して、他の用途に転用した場合でも、ソフトウェア・オブジェクトは文書データのどの部分を修正すればよいかを判断する情報を持たず、すべてユーザの判断に任されることになる。

従来の自然言語インターフェースは、以上のような従来のソフトウェア・オブジェクトが抱えるデータの再利用性の問題を解決していない。ユーザがソフトウェア・オブジェクトを使って既に作成したデータには、ある特定の用途に限られたものではあるが、ユーザの要求を実現するのに必要十分な情報が含まれている

。従って、これらのデータを他の用途に転用したい場合、他のソフトウェア・オブジェクトを使って表現形式を変更したり、既存のデータの一部だけを修正することで、望みのデータが得られことが多い。実際、人間は日常的に、既存のデータを元にして、その構造や意味を理解して、適切なソフトウェア・オブジェクトを選択し、最小限の修正で望みのデータを作成することを行っている。しかし、従来の自然言語インターフェースは、ユーザが入力する自然言語で表現された要求には着目しているが、ソフトウェア・オブジェクトが出力したデータを再利用することには着目していないため、ユーザに対して多大な負荷を強いる結果となっている。

本発明は以上のような課題を解決するために成されたものであり、その目的とするところは、複数の異なるソフトウェア・オブジェクトを統一的に操作できる汎用性と、ユーザの要求を実現するために作成したデータを他の用途に転用することができる柔軟性を備えた自然言語インターフェースを実現するための技術を提供することにある。

上記課題を解決するために成された本発明は、
特定のソフトウェア・オブジェクトで作成された既存のデータを指定する手順、
前記データを解析し、前記ソフトウェア・オブジェクトが該データを作成するために実行した命令記述列に変換する手順、
前記命令記述列をユーザに理解可能な自然言語による機能記述表現に変換する手順、及び、

前記機能記述表現から意味的な情報を獲得し、前記機能記述表現に意味を付加した要求記述表現を生成する手順
を含む処理をコンピュータに実行させることを特徴とする。

また、本発明は、コンピュータに上記のような処理を実行させるためのプログラムを提供する。

本発明による処理の手順は次の通りである。まず、ユーザが直接的に、もしくは

は間接的に、ユーザの要求を実現するために参照するデータを指定する。ここで言う「直接的」とは、ユーザ自身または他のユーザが特定のソフトウェア・オブジェクトを使って既に作成したデータ（一般にはデータファイル）を、コンピュータのOS（オペレーティング・システム）の機能を使って指定することである。また、「間接的」とは、ユーザがコンピュータの入力手段を使って入力した要求を解釈して、その要求を実現するのに最も適切なデータを既存のデータから選択し、指定することである。ここでは「映画鑑賞会のお知らせ」というタイトルの文書データが、ユーザにより直接的に指定されたものとする。

次に、指定したデータを解析して、そのデータを作成したソフトウェア・オブジェクトが実行可能な命令記述列に変換する。具体的には、データの各書式と、ソフトウェア・オブジェクトが実行可能な命令記述との対応を表した辞書（以下、データ解析部辞書と呼ぶ）を用いて、指定したデータを命令記述列に変換する。データ解析部辞書は、指定したデータを作成するソフトウェア・オブジェクトごとに用意する。前述の「映画鑑賞会のお知らせ」というタイトルの文書データが指定された場合は、文書データを作成した特定のワードプロセッサ用のデータ解析部辞書が選択され、例えば、文書データ内にサイズが24ポイントの文字列が見つかり、`Selection.Font.Size = 24` のように変換する。

次に、変換した命令記述列から、ユーザが理解可能な自然言語による機能記述表現に変換する。具体的には、各命令記述と、ユーザが理解可能な自然言語による機能記述表現との対応を表した辞書（以下、命令伝達部辞書と呼ぶ）を用いて、命令記述列を機能記述表現に変換する。命令伝達部辞書は、命令記述列を実行可能なソフトウェア・オブジェクトごとに用意する。前述の「映画鑑賞会のお知らせ」というタイトルの文書データが指定された場合は、文書データを作成した特定のワードプロセッサ用の命令伝達部辞書が選択され、例えば、「`Selection.Font.Size = 24`」という命令記述を、「選択した文字列の大きさを24ポイントにする」のように変換する。

次に、変換した機能記述表現から意味的な情報を獲得し、機能記述表現に意味を付加した要求記述表現を生成する。具体的には、各機能記述と、ユーザの要求の意味論的表現との対応を表した辞書（以下、機能翻訳部辞書と呼ぶ）を用いて、

機能記述表現から要求記述表現を生成する。機能翻訳部辞書の一例を第2図に示す。第2図の機能翻訳部辞書は、機能記述表現および要求記述表現に用いられる語句について、相互に置き換え可能な語句組（語句A、語句B）を一つの変換対とした変換表である。この変換表により、例えば「文字列のサイズを大きくする」という語句は、「文字列を視覚的に強調する」という語句に変換できる。また逆に、「文字列を視覚的に強調する」を「文字列のサイズを大きくする」に変換することも可能である。また、第2図の「(\$A=)」という書式は、語句内の可変要素を表す変数で、「(\$A=12)」のような形式で具体的な値を代入することで、通常の機能記述表現および要求記述表現として使える。

第2図の辞書に基づく機能記述表現から要求記述表現の生成手順を更に具体的に説明する。例えば、

「選択した文字列の大きさを24ポイントにする」

「選択した文字列をセンタリングする」

「選択した文字列を波型にする」

という一連の機能記述表現は、機能翻訳部辞書内にある、

「文字列のサイズを大きくする」×「文字列のサイズを(\$B=24)ポイントにする」

「文字列を視覚的に強調する」×「文字列のサイズを大きくする」

「文字列を視覚的に強調する」×「文字列をセンタリングする」

「文字列を視覚的に強調する」×「文字列を波型にする」

といった変換対に適合することから、

「選択した文字列は視覚的に強調されている」

という記述に変換することができる。

さらに、

「文字列『映画鑑賞会のお知らせ』を入力する」

という記述が機能記述表現の先頭にあると、機能翻訳部辞書内にある、

「文書のタイトル」×「『(\$A=)のお知らせ』」

「文書のタイトルを入力する」×「文字列『(\$A=)』を入力する/文字列を文書冒頭に配置する」

といった変換対に適合することから、

前述の「選択した文字列は視覚的に強調されている」という記述と併せることにより、

「文書のタイトルの文字列『(\$A1=(\$A2=映画鑑賞会)のお知らせ)』を入力する」という記述に変換することができる。すなわち、入力された『映画鑑賞会のお知らせ』という文字列は、文書のタイトルと認識されたことになる。

これらの結果より、

「文字列『映画鑑賞会のお知らせ』を入力する」

「入力した文字列を選択する」

「選択した文字列の大きさを24ポイントにする」

「選択した文字列をセンタリングする」

「選択した文字列を波型にする」

という一連の機能記述表現から、最終的に、

「文書のタイトルの文字列『(\$A1=(\$A2=映画鑑賞会)のお知らせ)』を入力する」

「文書のタイトルの文字列の大きさを(\$B=24)ポイントにする」

「文書のタイトルの文字列をセンタリングする」

「文書のタイトルの文字列を波型にする」

というテンプレート形式の要求記述表現が得られる。

以上のように、本発明は、ソフトウェア・オブジェクトにより作成された既存のデータから、ユーザの要求を実現するための自然言語による要求記述表現を、自動的に生成する基本的なアーキテクチャを提供するものである。このような本発明により、ソフトウェア・オブジェクトと自然言語インターフェースとの連携を従来よりも簡単に実現することができる。すなわち、生成された要求記述表現は特定のソフトウェア・オブジェクトに依存しない自然言語で表現されているため、もしユーザの要求を、他のソフトウェア・オブジェクトを操作することにより実現する場合でも、要求記述表現と命令記述列とを関連付ける辞書を作成するだけで、そのソフトウェア・オブジェクトを操作できる仕組みを実現することができる。

また、生成された要求記述表現は、ユーザが理解可能な自然言語で表現されて

いるということだけでなく、ユーザの要求に含まれる意味的な情報が加味され、かつ可変要素が認識されているため、要求の可変要素部分の最小限の変更だけで、他の用途に転用することができる。その結果、ユーザ自身または他のユーザが過去に作成した膨大なデータを、様々な用途に有効に利用することができる。

例えば、本発明の方法で得られる要求記述表現は自然言語表現であるから、それはそのまま特願2002-76319の方法における入力情報とすることができる。このように2つの発明を組み合わせることにより、自然言語を介したデータ変換が可能になる。例えば、Aというワードプロセッサで作成されたデータをBというワードプロセッサで利用したいが、BにはAの出力データを読み込むための変換フィルタがなく、しかも、Aというワードプロセッサが手元にない場合、まず、本発明に係る方法でAの出力データを自然言語の中間データに変換し、更にその中間データを特願2002-76319の方法でBの形式のデータに変換するのである。このとき、中間データの段階でユーザはその内容を自然言語の要求記述表現で確認することができ、必要ならそこで適宜内容を編集してからBに読み込ませることも可能である。要求記述表現には「タイトル」「宛名」「日付」というように、文書の構成要素の意味を示す情報が含まれているため、AやBの操作を全く知らないユーザでも容易に文書編集が可能である。

他にも、例えば、本発明の方法で得られる自然言語の要求記述表現から元のデータファイルの構造を解析することができる。また、本発明の方法で機器制御設定データを自然言語の要求記述表現に変換し、ユーザが適宜修正してから、特願2002-76319の方法で元の機器設定データに戻すことも可能である。また、本発明の方法でサンプル文書データを自然言語の要求記述表現に変換すれば、サンプル文書が出来上がるまでの操作手順を示す文書が出来上がる。このような文書は、例えばアプリケーション操作学習用テキストの作成に利用できる。また、本発明の方法で得られた自然言語の要求記述表現を既存の翻訳プログラムで他国語に翻訳すれば、単にテキスト情報だけでなく書式やレイアウト等に関する指示まで当該他国語に翻訳された文書が得られる。

図面の簡単な説明

第1図 本発明に係る方法によるデータ記述の生成と利用手順の一例を示すフローチャート。

第2図 機能翻訳部辞書の構成例。

第3図 本発明の一実施例であるコンピュータ・システムのハードウェア構成を示すブロック図。

第4図 本発明に従って構成された自然言語インターフェースの機能的構成を示すブロック図。

第5図 データ指定部辞書の一部の構成例。

発明を実施するために最良の形態

本発明の一実施例であるシステムについて図面を参照しながら説明する。本実施例のシステムは、特願2002-76319に記載の方法を応用してファイルを間接的に指定し、そのファイルのデータを本発明による方法で自然言語の要求記述表現に変換する。本実施例のシステムの概略構成を第3図に示す。このコンピュータ・システムは、一般のパーソナル・コンピュータを利用して構成されたものであって、CPU（中央演算処理装置）10、ROM（読出し専用主記憶装置）11、RAM（読書き可能主記憶装置）12、外部記憶装置コントローラ13と外部記憶装置（補助記憶装置）14、外部と連携するためのネットワーク・コントローラ15、ユーザ・インターフェース・アダプター16、ディスプレイ・コントローラ21とディスプレイ22、サウンド・コントローラ23、スピーカー24を備えている。ユーザ・インターフェース・アダプター16には、言葉列を入力するための各種入力機器（キーボード17、音声入力用マイク18、マウス19、手書き入力用タブレット20）が接続されている。

本実施例のシステムの機能的構成を第4図に示す。第4図において、自然言語入力部30は自然言語による言葉、言葉列又は文（以下、これらをまとめて「言葉」と呼ぶ）の入力を受け付け、その言葉を表す文字列を生成する手段である。言葉の入力方法としては、キーボード17を用いたキー入力、マイク18を用いた音声入力、マウス19を用いた画面上の文字入力パネルの操作、あるいは、タブレット20を用いた手書き入力から選択できる。もちろん、これら以外の方法

でも、そのための入力機器とそのためのソフトウェア（ドライバ）を用意さえすれば、言葉の入力に利用できる。

自然言語解析部 34 は、自然言語解析、辞書によるパーシング、対話形式構文生成、カテゴリ辞書マネージャ機能を有しており、上記文字列を解析して意味論的表現を生成する。文字列の解析には、自然言語処理の分野で一般的に知られた技術が利用できる。例えば、奈良先端科学技術大学院大学の「茶筌」、京都大学の「KNP」等の自然言語解析エンジンが既に知られているが、このような既存のエンジンを用いて自然言語解析部 34 を構成することができる。

データ指定部 36 は、意味論的表現に現れる全ての概念でデータ指定部辞書 42 を検索し、ユーザの要求に応じた処理を行うのに最適なソフトウェア・オブジェクト、およびそのソフトウェア・オブジェクトが作成したデータを選択する。データ指定部辞書 42 には、意味論的表現に用いられる概念とシステムで利用可能なソフトウェア・オブジェクト、およびそのソフトウェア・オブジェクトが生成したデータとを関連付ける情報が保存されている。

データ解析部 37 は、データ指定部 36 により指定したデータを構文的に解析し、指定したデータを作成したソフトウェア・オブジェクトが、指定したデータを作成するための実行可能な命令記述列に変換する。

命令記述変換部 38 は、データ解析部 37 により作成された命令記述列を、命令伝達部辞書 44 を検索することにより、自然言語で表現された機能記述表現に変換する。

機能記述変換部 40 は、命令記述変換部 38 により作成された機能記述表現を、機能翻訳部辞書 46 を検索することにより、機能記述表現から意味的な情報を獲得し、機能記述表現に意味を付加した要求記述表現を生成する。機能翻訳部辞書 46 の構成例は第 2 図に示した通りである。

機能翻訳部 41 は、意味論的表現に現れる全ての概念で機能翻訳部辞書 46 を検索し、辞書に記憶されているソフトウェア・オブジェクトの機能に合った機能記述表現に置き換える。

命令伝達部 39 は、機能翻訳部 41 により作成された機能記述表現に現れる全ての概念で命令伝達部辞書 44 を検索し、辞書に記憶されているソフトウェア・

オブジェクト 45 の機能を実行するための命令記述列を生成する。命令記述列は、例えば、該当ソフトウェア・オブジェクト 45 の API とパラメータ、あるいは、コマンドストリームを介して渡すコマンド列である。命令伝達部 39 は命令記述列を実行し、該当ソフトウェア・オブジェクト 45 の機能を実行する。

応答発生部 33 は、命令伝達部 39 により実行されたソフトウェア・オブジェクト 45 の実行結果を受け取り、ユーザが所望する形態で応答する。応答の形態は、例えば、ディスプレイ 22 への表示、プリンタ（図示せず）による印刷、データベースへの情報格納、機器の制御等、様々なものが考えられる。

辞書管理部 35 は、データ指定部辞書 42、データ解析部辞書 43、命令伝達部辞書 44、機能翻訳部辞書 46 の新規情報の作成、格納情報の変更、削除、閲覧表示を行う。さらに、制御部 32 は自然言語入力部 30、自然言語解析部 34、データ指定部 36、データ解析部 37、命令記述変換部 38、命令伝達部 39、機能記述変換部 40、機能翻訳部 41、応答発生部 33、ユーザ対話部 31、辞書管理部 35 に必要なデータを受け渡し、一連の動作を制御する。

「映画鑑賞会のお知らせを作成したい」という入力文字列の処理を本実施例のシステムで行う手順について第 1 図、第 2 図及び第 5 図を参照しながら説明する。

映画鑑賞会のお知らせを作成しようと考えたユーザが、例えばキーボード 17 を用いて「映画鑑賞会のお知らせを作成したい」という文を打ち込むと、自然言語入力部 30 が、キーボード入力インターフェースを通じて「映画鑑賞会のお知らせを作成したい」という文字列を受け取る（ステップ 50）。この文字列は自然言語解析部 34 に渡される。

自然言語解析部 34 は受け取った文字列を解析し、例えば「映画（の）」、「鑑賞会（の）」、「お知らせ（を）」及び「作成（したい）」という、構文および意味的に分離した 4 つの言葉から成る意味論的表現を生成する（ステップ 51）。この意味論的表現はデータ指定部 36 に渡される。

データ指定部 36 は、データ指定部辞書 42 に基づいて上記 4 つの言葉のソフトウェア・オブジェクト別評価点を取得する。データ指定部辞書 42 の一部の構成例を第 5 図に示す。第 5 図の辞書を用いた場合、「映画（の）」「鑑賞会（の）」

）」「お知らせ（を）」「作成（したい）」という意味論的表現から、

ワードプロセッサの評価点=1.7

電子メールクライアントの評価点=1.2

作図ソフトの評価点=0.1

というようにソフトウェア・オブジェクト毎の評価点が得られる。この結果から、データ指定部36は、総合的な評価点の最も高いソフトウェア・オブジェクトは「ワードプロセッサ」とであると判断し、データ指定部辞書42に格納されている「ワードプロセッサ」向けのデータ指定処理を実行する。

データ指定処理は、意味論的表現から適切なデータを検索し、評価することにより、最適なデータを指定する。例えば、意味論的表現に含まれる言葉をファイル名に含むデータファイルであって、先に選択したソフトウェア・オブジェクト（今の例では「ワードプロセッサ」）により作成されたデータファイルを検索し、最新の日付のデータファイルを最適と評価して指定する。なお、最適なデータファイルを自動評価する前に候補ファイルの一覧を表示し、ユーザに確認及び選択の機会を与えるようにしてもよい。ここでは「映画鑑賞会」という名前のデータファイルが指定されたとする。（ステップ52）

データ解析部37は、データ指定部36により指定したデータを構文的に解析し、指定したデータを作成するためのソフトウェア・オブジェクトが実行可能な命令記述列に変換する。例えば、データ指定部36で指定したデータがワードプロセッサにより作成された文書データであった場合、データ解析部37はそのワードプロセッサ用のデータ解析部辞書43により、文書データ内に格納されている文字列データ、および文字列の大きさや位置を指定したレイアウト情報を読み取り、ソフトウェア・オブジェクトが実行可能な命令記述に変換する。例えば、文書データ内にサイズが24ポイントの文字列が見つかり、「Selection. Font. Size = 24」のように変換する。（ステップ53）

次に、命令記述変換部38は、データ解析部37により作成された命令記述列を、命令伝達部辞書44を検索することにより、特定のソフトウェア・オブジェクトに依存しない、自然言語で表現された機能記述表現に変換する。例えば、「Selection. Font. Size = 24」という命令記述を、「選択した文字列の大きさを24

ポイントにする」のように変換する。(ステップ54)

機能記述変換部40は、命令記述変換部38により作成された機能記述表現を、機能翻訳部辞書46(第2図)を検索することにより、機能記述表現から意味的な情報を獲得し、機能記述表現に意味を付加した要求記述表現を生成する。例えば、

「文字列『映画鑑賞会のお知らせ』を入力する」

「入力した文字列を選択する」

「選択した文字列の大きさを24ポイントにする」

「選択した文字列をセンタリングする」

「選択した文字列を波型にする」

という一連の機能記述表現から、

「文書のタイトルの文字列『(\$A1=(\$A2=映画鑑賞会)のお知らせ)』を入力する」

「文書のタイトルの文字列の大きさを(\$B=24)ポイントにする」

「文書のタイトルの文字列をセンタリングする」

「文書のタイトルの文字列を波型にする」

という要求記述表現を生成する。(ステップ55)

次に、機能翻訳部41は、要求記述表現に含まれる可変要素を再定義して、機能記述表現に翻訳する。機能翻訳部41は、要求記述表現に含まれる可変要素を見つけると、ユーザ対話部31を介してユーザに可変要素部分の値を問い合わせる。ユーザが定義に相当する言葉(文字列)を入力すると、機能翻訳部41は要求記述表現の可変要素部分をその言葉で置き換え、機能記述表現に変換する。例えば、要求記述表現に、「日付の文字列『(\$Y=)年(\$M=)月(\$D=)日』を入力する」があれば、三つの可変要素の値をユーザに問い合わせ、
「日付の文字列『(\$Y=2002)年(\$M=12)月(\$D=1)日』を入力する」という機能記述表現に翻訳する。同様に、「場所の文字列『(\$P=)』を入力する」、「題名の文字列『(\$T=)』を入力する」、「あらすじの文字列『(\$S=)』を入力する」という要求記述表現の可変要素部分の値をユーザに問い合わせることにより、可変要素の含まれない機能記述表現に翻訳する。(ステップ56)

次に、命令伝達部39は可変要素を含まない機能記述表現を、命令伝達部辞書

44を用いて命令記述列を生成し、ソフトウェア・オブジェクト45に対してその命令を実行する。例えば、「ワードプロセッサを起動する」という機能記述表現に対しては、ハードディスクの所定場所に保存されたワードプロセッサのプログラムを読み出して実行するための命令記述列を生成し、OSに実行させる。また、「日付の文字列『(\$Y=2002)年(\$M=12)月(\$D=1)日』を入力する」という機能記述表現に対しては、文字列挿入機能を呼び出すための命令記述列を生成し、OSを通じてワードプロセッサのプログラムに実行させる。OSに渡すべき命令記述列はOSのアプリケーション・プログラミング・インターフェース(API)の仕様に従って作成し、ワードプロセッサのプログラムに渡すべき命令記述列はワードプロセッサのAPIの仕様に従って作成する。(ステップ57)

このようにして、ユーザはユーザ対話部31のガイダンスに沿い、「文書のタイトル」、「日付」、「場所」、「題名」、「あらすじ」という最小限の情報を入力することで、「映画鑑賞会のお知らせ」という文書データを簡単に作成できる。

以上、本発明の実施例について説明したが、実施例は上記に限られるものではない。例えば、作成したデータが案内文書で、文書内に「山田様」もしくは「会員各位」という文字列があった場合に、機能翻訳部辞書により「宛名の文字列『(\$N=山田)様』を入力する」もしくは「宛名の文字列『(\$N=会員)各位』を入力する」という要求記述表現が生成される。すなわち、案内文書の宛名部分に可変要素である変数が含まれていることが認識されることになる。ここで、本発明の実施例のシステムに、既存の会員名簿データベース管理システムを連動させると、宛名の可変要素である変数\$Nに、会員名簿データベース内の「会員名」を結び付けることにより、各会員宛ての大量の案内文書を一度に作成することもできる。

また、上記実施例ではワードプロセッサのようなソフトウェア・オブジェクトが作成したコンテンツ・データを対象としたが、他のソフトウェア・オブジェクトやハードウェア機器などを制御するための制御シーケンス・データにも有効である。例えば、あるハードウェア機器を操作するための制御シーケンス・データを自然言語で表現し、音声出力型コントローラを設けて読み上げることにより、機器操作に不慣れな初心者に対しても、より良い操作インタフェースを提供する

ことができる。

さらにまた、本発明では、ソフトウェア・オブジェクトが生成するデータを、その書式や内容に関わらず、ユーザの要求を実現するために用いられた自然言語で表現することができるため、自然言語文字列によるデータの検索効率を向上させることができる。例えば、「映画」という文字列をキーとしたデータ検索により、「映画鑑賞会のお知らせ」というタイトルの文書データだけでなく、従来の非言語的な表現形式では不可能であった、「映画番組を予約する」ためのビデオ機器の制御シーケンス・データなども探知することができ、データの再利用性をさらに向上させることができる。

請 求 の 範 囲

1. 特定のソフトウェア・オブジェクトで作成された既存のデータを指定する手順、

前記データを解析し、前記ソフトウェア・オブジェクトが該データを作成するために実行した命令記述列に変換する手順、

前記命令記述列をユーザに理解可能な自然言語による機能記述表現に変換する手順、及び、

前記機能記述表現から意味的な情報を獲得し、前記機能記述表現に意味を付加した要求記述表現を生成する手順、

を含む処理をコンピュータに実行させることを特徴とする自然言語による既存データの記述方法。

2. 請求項1に記載の方法をコンピュータに実行させるためのプログラム。

Fig. 1

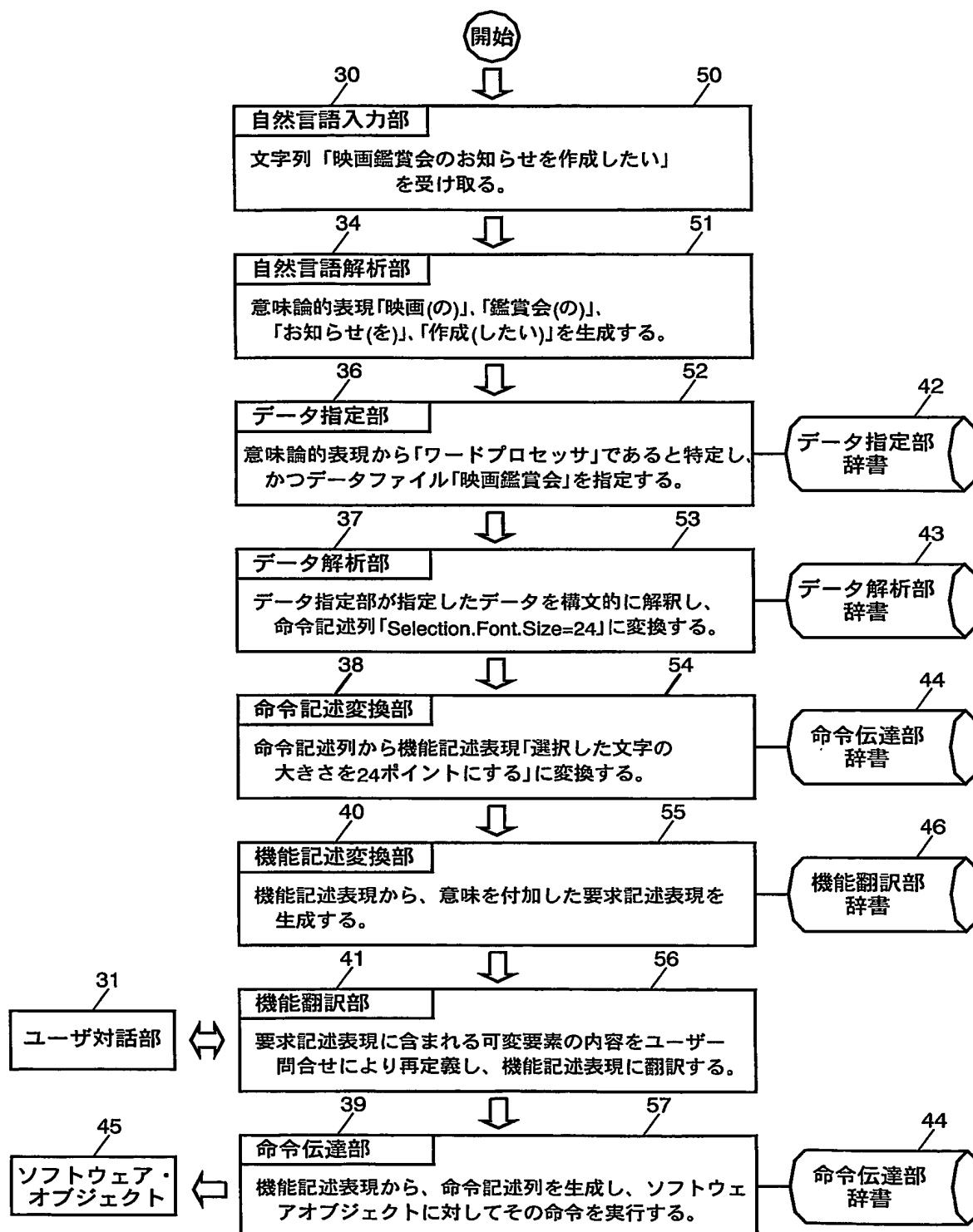


Fig. 2

語句A	種類	語句B
文字列を視覚的に強調する	目的語+動詞	文字列のサイズを大きくする
文字列を視覚的に強調する	目的語+動詞	文字列をセンタリングする
文字列を視覚的に強調する	目的語+動詞	文字列を波型にする
文字列を視覚的に強調する	目的語+動詞	文字列に色を付ける
文書のタイトルを入力する	目的語+動詞	文字列『(\$A=)』を入力する/文字列を文書冒頭に配置する
文書のタイトルを入力する	目的語+動詞	文字列『(\$A=)』を入力する/文字列を視覚的に強調する
文書の章を入力する	目的語+動詞	文字列『(\$A=)』を入力する/文字列を左寄せに配置する
日付を入力する	目的語+動詞	文字列『(\$A=)』を入力する/文字列を文書冒頭に配置する
宛名を入力する	目的語+動詞	文字列『(\$A=)』を入力する/文字列を文書冒頭に配置する
文書の章を入力する	目的語+動詞	文字列『(\$A=)』を入力する/文字列のサイズを大きくする
本文を入力する	目的語+動詞	文字列を入力する/文字列のサイズを標準にする
文字列のサイズを大きくする	目的語+動詞	文字列のサイズを(\$B=)ポイントにする
文字列に色を付ける	目的語+動詞	文字列の色を(\$C=)にする
文字列のサイズを標準にする	目的語+動詞	文字列のサイズを(\$B=10)ポイントにする
文書のタイトル	名詞	『(\$A=)のお知らせ』
文書のタイトル	名詞	『(\$A=)のご案内』
文書のタイトル	名詞	『(\$A=)について』
日付	名詞	『(\$Y=)年(\$M=)月(\$D=)日』
宛名	名詞	『(\$N=)様』
宛名	名詞	『(\$N=)各位』
:	:	:

Fig. 3

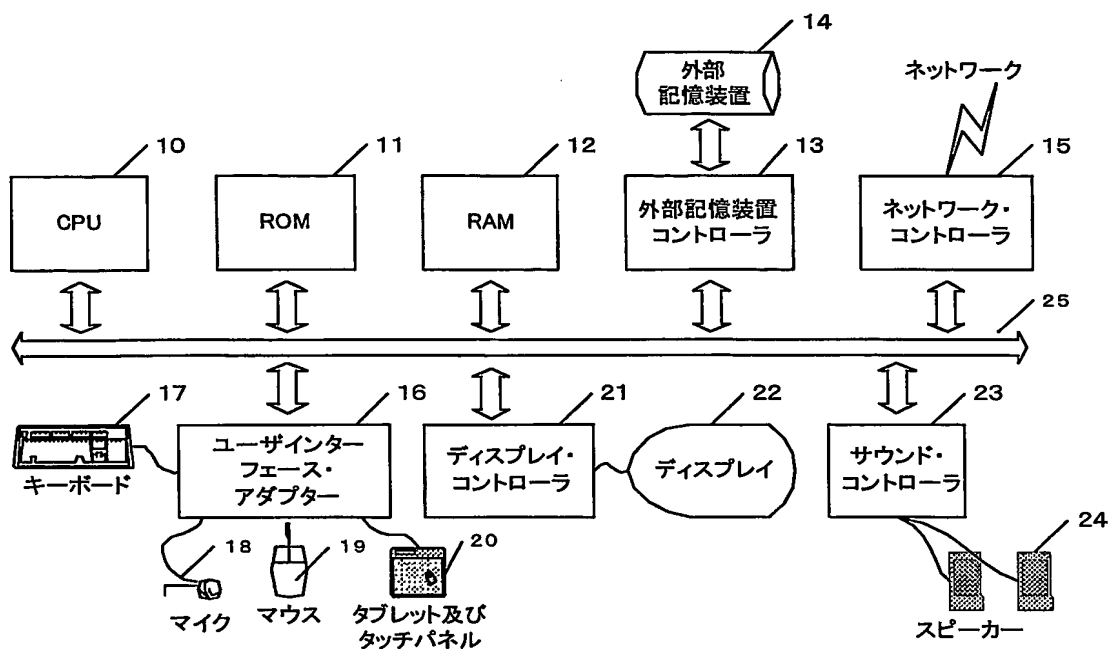


Fig. 4

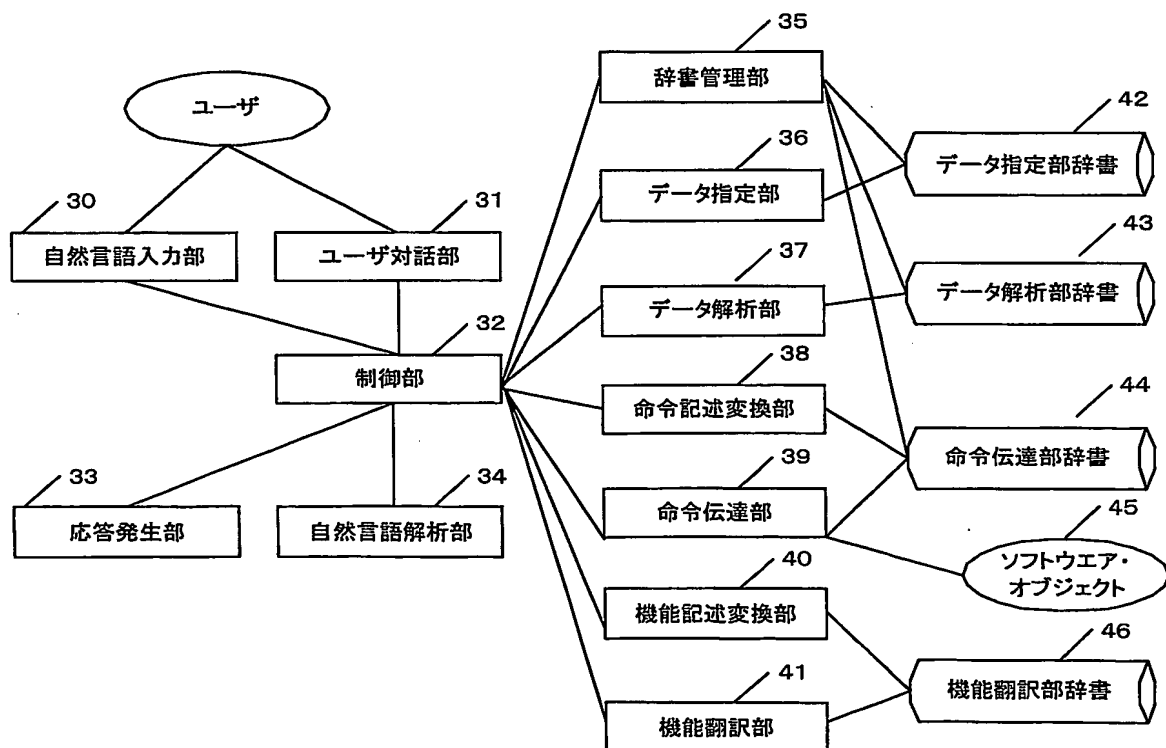


Fig. 5

入力語	種類	評価点	ソフトウェア・オブジェクト
:			
お知らせ	名詞	1	ワードプロセッサ
お知らせ	名詞	1	メール
:			
案内	名詞	0.6	ワードプロセッサ
案内	名詞	0.4	メール
:			
作成	動詞	0.7	ワードプロセッサ
作成	動詞	0.1	図形作成ツール
作成	動詞	0.2	メール
:			
送る	動詞	1	メール
:			

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/02278

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl.⁷ G06F3/00, G06F17/27, G10L13/08

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl.⁷ G06F3/00, G06F9/06, G06F9/44, G06F17/21, G06F17/27,
G10L13/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Toroku Jitsuyo Shinan Koho	1994-2003
Kokai Jitsuyo Shinan Koho	1971-2003	Jitsuyo Shinan Toroku Koho	1996-2003

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 10-133907 A (NEC Yonezawa, Ltd.), 22 May, 1998 (22.05.98), Full text; all drawings (Family: none)	1, 2
A	JP 6-083669 A (NEC IC Miconsystem Kabushiki Kaisha), 25 March, 1994 (25.03.94), Full text; all drawings (Family: none)	1, 2
A	JP 63-108454 A (Canon Inc.), 13 May, 1988 (13.05.88), Page 3, upper left column, line 12 to lower left column, line 5; Fig. 7 (Family: none)	1, 2

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not
considered to be of particular relevance"E" earlier document but published on or after the international filing
date"L" document which may throw doubts on priority claim(s) or which is
cited to establish the publication date of another citation or other
special reason (as specified)"O" document referring to an oral disclosure, use, exhibition or other
means"P" document published prior to the international filing date but later
than the priority date claimed

"T"

later document published after the international filing date or
priority date and not in conflict with the application but cited to
understand the principle or theory underlying the invention
document of particular relevance; the claimed invention cannot be
considered novel or cannot be considered to involve an inventive
step when the document is taken alone

"X"

"Y"

document of particular relevance; the claimed invention cannot be
considered to involve an inventive step when the document is
combined with one or more other such documents, such
combination being obvious to a person skilled in the art
document member of the same patent family

"&"

Date of the actual completion of the international search
03 June, 2003 (03.06.03)Date of mailing of the international search report
17 June, 2003 (17.06.03)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP03/02278

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 5-046656 A (Toshiba Corp.), 26 February, 1993 (26.02.93), Full text; all drawings (Family: none)	1, 2

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. C1' G06F 3/00, G06F 17/27, G10L 13/08

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. C1' G06F 3/00, G06F 9/06, G06F 9/44, G06F17/21,
G06F17/27, G10L13/08

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1922-1996年
 日本国公開実用新案公報 1971-2003年
 日本国登録実用新案公報 1994-2003年
 日本国実用新案登録公報 1996-2003年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP 10-133907 A (米沢日本電気株式会社) 1998.05.22 全文 全図 (ファミリー無し)	1, 2
A	JP 6-083669 A (日本電気アイシーマイコンシステム株式会社) 1994.03.25 全文 全図 (ファミリー無し)	1, 2
A	JP 63-108454 A (キヤノン株式会社) 1988.05.13 第3頁左上欄12行~同左下欄5行 第7図 (ファミリー無し)	1, 2

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」 口頭による開示、使用、展示等に言及する文献
 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」 同一パテントファミリー文献

国際調査を完了した日

03.06.03

国際調査報告の発送日

17.06.03

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)
 郵便番号100-8915
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)
 圓道 浩史



5E 3052

電話番号 03-3581-1101 内線 3520

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	J P 5 - 0 4 6 6 5 6 A (株式会社東芝) 1 9 9 3 . 0 2 . 2 6 全文 全図 (ファミリー無し)	1, 2

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.